DOCUMENT RESUME

ED 403 744 FL 024 362

AUTHOR O'Sullivan, Barry; Porter, Don
TITLE Speech Style, Gender, and Oral Proficiency Interview Performance.
PUB DATE 96
NOTE 14p.; Paper presented at the Annual Meeting of the Southeast Asian Ministers of Education Organization Regional Language Center Seminar (31st, Singapore, 1996).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; English; *English (Second Language); Foreign Countries; Higher Education; Interviews; *Japanese; *Language Patterns; *Language Styles; *Language Tests; Language Usage; Native Speakers; Oral Language; *Sex Differences; Speech Skills; Testing
IDENTIFIERS *Oral Proficiency Testing

ABSTRACT

This study investigated: (1) whether there is a gender effect in the speech of Japanese learners of English as a Second Language; (2) whether the effect is positive if the interlocutor is female; (3) whether there are associations of gender effect with specific features of speech; and (4) in which linguistic features of learner speech the gender effect is most evident. Subjects were six female and six male Japanese university students, average age approximately 20 years, and observed by three male and three female native speakers of English. Each Japanese student was interviewed twice, once by a man and once by a woman, and observed by a native English speaker of the same gender. Interviews were analyzed for specific speech characteristics (e.g., use of fillers, rephrasing, minimal response, repetition) and scored for accent, grammar, vocabulary, fluency, and comprehension. Results for the 24 interviews indicate a significant difference in scores awarded by different interviewer/observer pairs. Despite a high degree of agreement within pairs, and consistency in interviewer speech styles, in all but one interview the students scored higher when interviewed by a woman, particularly in grammar and fluency. However, male and female interviewers did show different patterns in linguistic features. Contains 15 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Speech Style, Gender and Oral Proficiency Interview Performance

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Barry O'Sullivan
Okayama University, Japan

Don Porter
University of Reading, England

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Barry O'Sullivan
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 403 744

Introduction

In the field of foreign language testing there is a steadily growing interest in those factors which affect the test performance of the testee (see Bachman, 1990, for a discussion of "test method facets"). Some of this interest is motivated by a desire to detect and then eliminate test features which are seen as distorting the tester's attempts to achieve accurate assessment of learners' language proficiency: these features are thus seen as sources of measurement error (Bachman et al. 1995, Kunnan 1995).

A number of researchers, however, distinguish between test features which are indeed irrelevant to the ability which is being measured, and those which are relevant to that ability (Locke 1984, Porter 1991, Porter and Shen 1991, O'Sullivan 1995, O'Sullivan and Porter 1995). If a feature affects test results to a significant degree, but is irrelevant to the ability being measured, it is indeed a source of measurement error which needs to be eliminated. If it is relevant to the ability being measured, however, and occurs in tests because it is an essential and naturally occurring part of natural language use, and if it affects test results to a significant degree, it is desirable—in fact necessary—that it should be included in test activities and tasks. Such features should be seen as contributing to test validity, whereas the former features should be seen as detracting from test validity. It is then an important goal of research related to language testing to

FL023467

discover which test features constitute significant sources of error in learners' performance. It goes without saying, perhaps, that test features which do not have a significant effect on learners' performance are irrelevant to the task in hand and can be ignored.

One particular feature which has been fairly consistently shown to affect learners' performance on tests of spoken interaction to a significant degree is the gender of the person with whom the person interacts (Locke 1984, Porter and Shen 1991, Porter 1991). Henceforth we shall use the term *gender effect* to refer to variation in linguistic features of learners' language which can be systematically related to differences in the gender of interlocutors. Gender effects have been found in the spoken interaction of learners from varied cultural backgrounds, although there is some evidence that the nature of such effects may vary with the cultural background of the learner. Thus while it has generally been found that the spoken language of learners in interviews will be rated more highly by independent raters when the interviewer is a woman, a small number of studies suggest that Arab speakers of English tend to achieve higher independent ratings when they are interviewed by a man (Locke 1984, Porter 1991).

Although the evidence so far on the basis of research done with Latin American, North African, and Middle Eastern language learners suggest that the gender of an interlocutor *may* produce significant effects in the spoken language production of learners from all cultures, it has yet to be shown that interlocutor-gender is indeed a systematic and significant factor affecting the quality of spoken foreign language produced by Asian—specifically Japanese—learners. Moreover, if such an effect is found, it is a matter of some interest to discover whether the spoken foreign language of

Japanese learners is positively or negatively affected by the gender of their interlocutors.

Finally, it has been suggested that the superior quality of spoken language produced by learners from many cultures when the interlocutor is a woman, may result from specific features to be found in the distinctive ways, often characterised as 'supportive' (see for example Coates 1993), in which women of many cultures use language (Fishman 1978, Wolfson 1989). It is thus important and of considerable interest to discover what the critical and distinctive features of women's speech are.

The study reported in this paper proposes to shed light , then, on four research questions:

1. Is there evidence of a gender effect in Japanese learners' spoken English?

2. If there *is* a gender effect in the spoken English of Japanese learners, is it positive when the interlocutor is female?

3. Where a gender effect is noted, can this be systematically associated with specific features of the interlocutor's speech?

4. Where a gender effect is noted, in which linguistic features of learners' speech is it made most evident?

The Study

The subjects involved in this study included six female and six male Japanese university students, average age approximately 20 years, and six native speakers of English, three female and three male, average age 29.6 years.

Each Japanese student was interviewed and observed by some of the native speakers. All interviews were conducted under similar conditions in a pair of adjacent interview rooms at the Department of Applied Psychology, Okayama University, Japan, over a two week period in November 1995. The interviews were video taped and audio taped in case of occasional lack of clarity in the audio tape. The interview format was structured, with two parts, the first part being designed to elicit short answers, while in the second part the subject was encouraged to produce longer responses. This interview type is similar to that employed by O'Sullivan (1995).

Subjects were interviewed twice, once by a woman and once by a man. On both occasions an observer of the same gender as the interviewer was also present. The requirement that the interviewer and observer be of the same gender was intended to ensure that any gender effect should not be compromised by an effect resulting from a different gender in the observer. The interview schedule was balanced to control for an order effect, by ensuring that half of the candidates—comprised of an equal number of women and men— were first interviewed by women while the remainder were first interviewed by men. The performance was scored at the time of the interview by both the interviewer and the observer, using the analytic rating scale developed by the American Foreign Service Institute.

For the first section of the study, investigating any possible interviewer/observer-gender effect, the scores obtained by these raters were analysed using a two factor ANOVA. In addition, interviewer and observer scores were compared using the Spearman rho statistic in order to establish inter-rater reliability.

The second area of interest to this study concerned the language used by the interviewer in the interaction. The interviews which proved most

interesting, in terms of variation of scores awarded, were transcribed and analysed using the framework described below. The results were tabulated and frequency of item occurrence used to identify the different speech characteristics of the women and men interviewers. Once established, the scores achieved by subjects when interviewed under both conditions were compared using t-tests.

Transcripts of learners' spoken interaction were examined for a number of speech style characteristics, similar to, though somewhat more extensive than those employed by Porter and Shen (1991). The characteristics examined were selected in order to investigate the language of the interviewers in terms of their question and response type. The categories were:

Question	**Example**
Fillers (F)	This includes the use of such fillers as 'well', 'uh', 'OK', 'um', etc.
Rephrasing (RP)	Interviewer paraphrases the candidate's response
Repetition (R)	Interviewer repeats own utterance
Question Refocus (QR)	No response time given to candidate, interviewer immediately rephrases or redirects the question.
Other (O)	

Response	
Minimal Responses (MR)	The interviewer responds to a candidate with utterances such as 'yeah', 'mmmm', 'uh-huh' etc.
Repetition (R)	Interviewer repeats the candidate's utterance
Clarification Requests (CR)	Where the interviewer explicitly requests a clarification by the candidate of an utterance made by the candidate
Expansion (E)	These are questions/statements designed to elicit message expansion which deviate from the set question prompts, for example, "So what did you do after that?"
Expressions of Interest (EI)	Where the interviewer uses a phrase such as "Is that right?" or "That's interesting." or uses intonation to show a marked interest in the candidate's response.
Correction (C)	The interviewer uses one of three types of correction: Lexical usage; Pronunciation; Grammar
Other (O)	

The decision to employ the above framework was based on a preliminary study involving the examination of six interviews in which the same format had been used. While most of the elements of the framework were derived from the sociolinguistics literature (Zimmerman and West 1975, Brown and Levinson 1978, Fishman 1978, Maltz and Borker 1982), the element 'Question refocus' had not been previously referred to. The decision to include this element was due to the fact that it was observed on a significant number of occasions in the pilot study and so was added to the framework for this research.

Results

As this study is primarily interested in the effect on performance, as measured using an analytic measuring scale, of the gender of interviewer/observer partnership, it is to those scores that we now turn.

Analyses of Scores Awarded

Comparison of the scores awarded by the interviewers and observers (see Table 1) indicates a high degree of agreement in the scores awarded in the twenty four interviews.

Scale Element	Mead Diff.	DF	t-Value	P-Value	Sig.
Accent	-.083	23	-1.000	.3277	NS
Grammar	0	23	0	-	NS
Vocabulary	-1.083	23	-1.919	.0674	NS
Fluency	.083	23	.310	.7592	NS
Comprehension	-.396	23	-.528	.6027	NS
Overall Score	**-1.479**	**23**	**-.768**	**.4503**	**NS**

Table 1: Comparison of Interviewer and Observer Scores

As can be seen from the table, there is no element of the scale in which there is a significant difference between the scores awarded. It was not necessary to implement the fall back position of a third rating. The calculated Spearman rho coefficient of .746 indicates that the relationship between the two sets of scores is significant ($p<.05$) and substantial.

Results for the twenty four interviews indicate that there is a significant difference in the scores awarded by the different interviewer/observer partnerships. While the results seen in Table 1 indicate that there is a high degree of agreement within the pairs, in all but one of the interviews, the first one, the students scored higher when interviewed by a woman. The ANOVA carried out on the results (see Table 2) confirms that this observation is actually statistically significant ($p<.05$). Also of interest in Table 2 is the fact that the gender of the candidate does not appear to be a significant factor.

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Subject Gender	1	202.711	202.711	2.205	.1532
Interviewer	1	452.836	452.836	4.925	* .0382
Subject Gender x Interviewer	1	1.628	1.628	.018	.8955
Residual	20	1839.010	91.951		

* Significant ($p<.05$)

Table 2: ANOVA of test performance

When a similar ANOVA was carried out on the results awarded on all elements of the Analytic scale used (Table 3) it was observed that the areas in which a difference was found to be of significance were those of Grammar and Fluency. The weighting on the individual elements of the scale used means that these two provide approximately 48% of the available marks, so the large difference seen in the scores awarded,

especially for Grammar, can be seen as the principal reason behind the overall significance.

	Accent	Grammar	Vocabulary	Fluency	Comprehension
Subject Gender	NS	NS	NS	NS	NS
Interviewer	NS	p<.005	NS	p<.05	NS
Subject Gender x Interviewer	NS	NS	NS	NS	NS

Table 3: Summary of ANOVA results on Elements of Analytic Scale

Analyses of Language Characteristics

While it was observed above that the scores awarded in eleven of the twelve cases when the interlocutor was female were higher than when the interlocutor was male, in eight of the cases this difference was found to be in the region of 10% or more. A descriptive analysis was performed on the transcripts of the sixteen interviews involved, the results of which are shown in Tables 4 and 5.

#	Testee	Tester	F	RP	R	QR	O	MR	R	CR	E	EI	C	O	Length
2	Man	Man 1	19	2	2	1	0	30	14	2	2	1	0	0	504
4	Man	Man 2	16	6	5	0	0	25	0	0	3	0	0	0	605
5	Man	Man 3	16	2	0	0	0	37	8	1	0	2	0	0	581
6	Man	Man 3	13	3	0	0	0	38	5	3	3	0	0	0	463
7	Woman	Man 1	21	6	1	0	0	34	2	6	4	1	0	0	557
9	Woman	Man 1	16	2	1	0	0	53	7	2	10	1	0	0	372
10	Woman	Man 2	17	8	3	0	0	24	1	1	4	2	1	0	544
11	Woman	Man 2	16	5	0	0	0	24	1	1	2	0	0	0	340

Table 4: Transcript Analysis for Men Testers

#	Testee	Tester	F	RP	R	QR	O	MR	R	CR	E	EI	C	O	Length
2	Man	Woman 1	7	0	3	0	3	16	1	2	2	4	0	0	524
4	Man	Woman 2	9	1	0	0	0	13	3	1	5	2	0	0	540
5	Man	Woman 2	4	0	0	2	0	39	3	1	11	5	0	0	798
6	Man	Woman 3	8	1	2	0	0	19	2	1	12	1	0	0	436
7	Woman	Woman 1	8	3	2	0	7	7	1	2	4	4	1	0	640
9	Woman	Woman 1	7	2	2	1	1	22	1	0	3	7	0	0	256
10	Woman	Woman 2	10	4	2	0	0	15	6	4	7	4	1	0	406
11	Woman	Woman 3	10	5	2	1	0	9	5	0	7	1	0	0	354

Table 5: Transcript Analysis for Women Testers

Further analysis of these results indicates that there is little observable overall difference between the speech styles of the interviewers (see Table 6). In order to made comparisons both with the results of Porter and Shen (1991) and between the different interviews in this study, these numbers in the table have been calculated as representing the number of occurrences per two minutes.

While there is a significant difference in the use of fillers, with the men producing almost twice as many on average, it does not appear to be an element of speech style which greatly affects the language of a communicative exchange, though its interaction with other aspects of speech style may well be important.

	F	RP	R	QR	O	MR	R	CR	E	EI	C	O	Length
M	4.1817	1.0405	0.3335	0.0298	0	8.4723	1.193	0.4964	0.9405	0.2037	0.0275	0	495.75
W	2.1916	0.6093	0.4773	0.1385	0.3085	4.5945	0.7527	0.3328	1.6405	0.9925	0.0604	0	494.25
t	0.0009	0.1585	0.4212	0.1727	0.0941	0.0348	0.3641	0.4202	0.1613	0.041	0.5163	-	-
Sig	Sig.	NS	NS	NS	NS	Sig.	NS	NS	NS	Sig.	NS	-	NS

Table 6: Average Occurrence per 2 Minutes of Interaction, with t-test results

What is interesting is the greater production of minimal responses by the men interviewers, again almost twice that of the women interviewers. Here the opposite situation would have been expected, as was the case with Porter and Shen (1991). It may be interesting to examine the different ways in which the men and women interviewers use minimal responses. Here for example a survey of the transcripts indicated a degree of intonational differences, with women employing a greater number of what may be described as bi-tonal or multi-tonal responses compared to the men's use of more mono-tonal responses which, to the testee, may sound more mechanical and therefore less supportive. Of course, the tendency for the men to employ a relatively lower pitched response than the women,

particularly when coupled with the above intonational differences, may be a more convincing explanation for an impression of lack of support for the learner.

The only areas in which the women were more productive than the men were in the use of expansion questions and in overtly expressing interest in what the candidates had to say ($p<.05$). In the case of the former it must be said that the numbers are really too small to draw any certain conclusions from, and while the difference was almost double it was not statistically significant. Of the latter it can be said that these figures clearly indicate the supportive nature of the women interviewers' speech style when compared to that of the men. Indeed all three of the men failed to use any expression of interest in at least one of their interviews, while this never occurred with any of the women. This appears to reinforce the observation of Brown and Levinson (1978) that women tend to express politeness and support by acknowledging and building on the utterances of other speakers in an interaction.

Conclusions

The results of the ANOVA performed on the scores awarded by the interviewer-observer pairs of different gender clearly indicated that, when this pair was made up of women, the candidate was more likely to achieve a higher score. This was true both in terms of the overall score awarded, on the most influential element of the scale, Grammar, and for Fluency. This result supports the findings of Porter and Shen (1991) and appears to establish the veracity of the claim that the gender of the interviewer/tester is indeed a factor which must be controlled for in any testing situation.

While this study focused on an examination of the scores achieved by women and men interviewees when involved in a language testing situation, it also undertook a brief examination of the actual language produced in those interactions. Preliminary analysis of this language established differences in the questions and responses of the women and men. While the men provided significantly more occurrences of both fillers and minimal responses, it was observed that there may be a question mark over the effect of the way in which they employed these two characteristics. Women, on the other hand, tended to show their support in a more emphatic way. It may well be that in using more expansion questions—defined here as questions which are a product of the tester's reaction to a testee's statement which has not been specifically suggested in the interview outline—and more especially in expressing interest in the responses of candidates more regularly, the women interviewers are changing the nature of the interview. If these actions elicit language which is more fluent, grammatically complex and/or accurate then this enhancement of testee performance may account for the higher scores awarded. Specific emphasis on these points in pre-interview training may be called for, in order to limit their effect.

Of real interest to this research is the extent to which the differences identified by the testers as being significant, in the areas of Grammar and Fluency, are to be found in the actual language produced by the candidates. In order to achieve this a more detailed analysis of the transcripts of these interviews might include an examination of the fluency, complexity and accuracy of the language (see for example Skehan and Foster 1995).

Finally, the results of this study, when seen in light of those earlier studies referred to in the introduction, highlight the relevance of continued

research in this area. The importance of this approach both to our ability to construct more valid tests and to our knowledge of language in use is clear.

References

Bachman, L. 1990. *Fundamental Considerations in Language Testing.* Oxford: OUP.

Bachman, L. F., F. Davidson, K. Ryan and I-C. Choi. 1995. *An investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL Comparability Study.* Cambridge: Cambridge University Press.

Brown, P. and S. Levinson. 1978. Universals in language usage: politeness phenomena. In E. Goody (Ed.) Questions and Politeness: Strategies in Social Interaction. Cambridge: Cambridge University Press. 56 - 310.

Coates, J. 1993. *Women, Men and Language.* Second edition. Harlow: Longman. First edition 1986.

Fishman, P. M. 1978. Interaction: The work women do. *Social Problems* 24: 397 - 406.

Kunnan, A. J. 1995. *Test Taker characteristics and test performance: A structural modeling approach.* Cambridge: Cambridge University Press.

Locke, C. 1984. Unpublished MA project, University of Reading.

Maltz, D. N. and R. A. Borker. 1982. A cultural approach to male-female miscommunication. In J. Gumperz (Ed.) *Language and Social Identity.* Cambridge: Cambridge University Press.

O'Sullivan, B. 1995. *Oral Language Testing: Does the Age of the Interlocutor make a Difference?* Unpublished MA Dissertation, University of Reading.

O'Sullivan, B. and D. Porter. 1995. *The importance of audience age for learner-speakers and learner-writers from different cultural*

backgrounds. Paper presented at the RELC Conference, Singapore. April 1995.

Porter, D. 1991. Affective factors in language testing. In C. J. Alderson and B. North (Eds.) *Language Testing in the 1990s*. Modern English Publications. Also published in *Review of English Language Teaching* 1 (1).

Porter, D. and Shen Shu Hung. 1991. Sex, Status and Style in the Interview. *The Dolphin 21*. Aarhus: Aarhus University Press. 117 - 128.

Skehan, P. and P. Foster. 1995. Task type and task processing conditions as influences on foreign language performance. *Thames Valley University Working Papers in ELT*. Volume 3: 139 - 188.

Wolfson, N. 1989. *Perspectives: Sociolinguistics and TESOL*. New York: Newbury House.

Zimmerman, D. H. and C. West. 1975. Sex roles, interruptions and silences in conversation. In B. Thorne and N. Henley (Eds.) *Language and Sex: Difference and Dominance*. Rowley, Mass.: Newbury House.